標準単体上の最小2乗問題に対する対数正則化と近接分離法 (数理最適化の発展 : モデル化とアルゴリズム)

AUTHOR(S):

田中, 未来; 武田, 朗子

CITATION:

田中, 未来 ...[et al]. 標準単体上の最小2乗問題に対する対数正則化と近接分離法 (数理最適化の発展 : モデル化とアルゴリズム). 数理解析研究所講究録 2018, 2069: 11-22

ISSUE DATE:

2018-04

URL:

http://hdl.handle.net/2433/241962

数理解析研究所講究録
第2069巻 2018年 11-22

標準単体上の最小 2 乗問題に対する対数正則化と近接分離法

田中未来 *　　武田朗子 †

概要

本論文では標準単体上における ℓ_2 損失関数と正則化項の和の最小化問題を考える. この問題に対する正則化項として, よく用いられる ℓ_1 正則化や ℓ_2 を用いることはある意味で不自然であるため, 本論文では対数正則化を用いることを提案する. 対数正則化を施した問題は Dirichlet 分布を事前分布とする最大事後確率推定問題として解釈できるほか, Kullback–Leibler 擬距離を罰則化した問題としても解釈できる. さらに本論文ではこの問題を解くための 3 つの近接分離法 (加速近接勾配法, 交互方向乗数法, 線形化交互方向乗数法) を提案し, それぞれのアルゴリズムの子問題を効率よく解くことができることを示す.

1 はじめに

標準単体上における最小 2 乗問題は多くの応用分野から現れる問題である. 後述するように, 材料科学における混合比の推定 (Tanaka et al., 2017) やハイパースペクトル画像の解析 (Heinz and Chang, 2001; Bioucas-Dias and Figueiredo, 2010; Heylen et al., 2011, 2013; Chouzenoux et al., 2014; Condat, 2017) などが応用例として挙げられる.

最小 2 乗問題 $\min(1/2)\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2$ はしばしば多重共線性の問題をもつ. すなわち, 計画行列 $\boldsymbol{A}$ が悪条件であるため, 最適解 $(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{b}$ が不安定になるということを意味する. この問題は標準単体上における最小 2 乗問題でも同様に発生する. この問題を防ぐために正則化が使われる. 本論文では, 標準単体上で ℓ_2 損失関数と正則化項の和を最小化する次のような問題を考える:

$$\begin{array}{ll}\text{minimize} & \dfrac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+r(\boldsymbol{x}) \\ \text{subject to} & \mathbf{1}^\top\boldsymbol{x}=1, \boldsymbol{x}\geq\mathbf{0},\end{array} \tag{1}$$

ここで, $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ と $\boldsymbol{b}\in\mathbb{R}^m$ は定数, $\boldsymbol{x}\in\mathbb{R}^n$ は決定変数, $\mathbf{1}\in\mathbb{R}^n$ はすべての要素が 1 のベクトル, $r:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ は正則化項である. 問題 (1) における正則化項 r としてなにを用いるかを考えることは本論文の主たる目的の 1 つである.

Lasso (Tibshirani, 1994) などにおいてスパースな解を得るためによく用いられる正則化に ℓ_1 正則化 $r(\boldsymbol{x})=\gamma\|\boldsymbol{x}\|_1$ がある. しかしながら, ℓ_1 正則化は問題 (1) に対しては意味がない. それは, $\boldsymbol{x}$ が標準単体上の点であるときは常に $r(\boldsymbol{x})=\gamma$ となり, 定数を足しているにすぎないためである.

リッジ回帰などにおいて悪条件な問題を良条件な問題にするためによく用いられる正則化に ℓ_2 正則化 $r(\boldsymbol{x})=(\gamma/2)\|\boldsymbol{x}\|_2^2$ がある. ℓ_2 正則化を施すことにより, 問題 (1) の目的関数の条件数が改善されるため, 最適解が安定化することが期待できる. ℓ_2 正則化を施した問題は箱型制約と 1 本の線形制約をもつ凸 2 次最適化問題なので, Dai and Fletcher (2006); Han et al. (2013) が提案した勾配法に基づくアルゴリズムを用

* 統計数理研究所 数理・推論研究系
† 統計数理研究所 数理・推論研究系, 理化学研究所 革新知能統合研究センター

いて解くことができる. ハイパースペクトル画像の解析の文脈では, いくつかの研究において ℓ_2 正則化が導入されている (Settle and Drake, 1993; Li and Du, 2015). しかしながら, 制約条件の一部またはすべてを無視して問題を解くなどしており, その扱いは十分であるとはいえない. その理由の 1 つは, 最適化アルゴリズムが存在するものの, 応用分野の研究者がすぐに使えるようなソフトウェアが存在しないためかもしれない. Chouzenoux et al. (2014) は正則化を施さない標準単体上での最小 2 乗問題に対する内点法を提案している. 彼らは正則化の重要性も指摘しており, 彼らの手法が正則化付きの問題にも拡張できると述べている. しかしながら, 具体的な定式化やアルゴリズムについては述べられておらず, 計算機実験の結果も存在しない. Tanaka et al. (2017) は ℓ_2 正則化付きの問題を商用ソフトウェアを用いて解いている.

ℓ_1 正則化と ℓ_2 正則化は制約のない最小 2 乗問題でよく使われる正則化項だが, 後述するように Bayes 的解釈に基づいて考察すると問題 (1) に対しては不自然であるといえる. 具体的には, ℓ_1 あるいは ℓ_2 正則化を施した問題を最大事後確率推定問題とみなしたときに, 対応する事前分布の台が実行可能領域と整合的でない.

本論文では, 問題 (1) の制約条件と整合的な正則化を提案する. 提案する正則化は $r(\boldsymbol{x}) = -\sum_{j=1}^{n} \gamma_j \log x_j$ であり, これを本論文では対数正則化と呼ぶ. ここで, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^\top$ は正のパラメータを並べたベクトルである. 対数正則化は多重共線性の問題を回避することができる. 具体的には, 問題 (1) の目的関数の条件数を向上させ, 最適解を安定させることができる. さらにこの正則化は Bayes 的な解釈をもつことを示すことができる. 対数正則化問題は Dirichlet 分布を事前分布とする最大事後確率 (MAP) 推定問題に対応しており, 事前分布の台が問題 (1) の制約条件と整合的である. さらに, 対数正則化は Kullback–Leibler (KL) 擬距離に基づく罰則ともみなすことができる. これらの解釈は ℓ_1 正則化や ℓ_2 正則化が持たないものである.

対数正則化は Weston et al. (2003); Candés et al. (2008); Mazumder et al. (2011) などによって研究され, Larsson and Ugander (2011) によって混合モデルに応用されている. しかしながら, これらの文脈で用いられている対数正則化は $r(\boldsymbol{x}) = \sum_{j=1}^{n} \gamma_j \log(x_j + \xi)$ という正則化である. ここで, $\xi > 0$ は目的関数を有界にするための微小な定数である. これはわれわれが考える対数正則化 $r(\boldsymbol{x}) = -\sum_{j=1}^{n} \gamma_j \log x_j$ とパラメータ ξ の存在を除いて逆符号であり, 性質は大きく異なる. 前述の既存研究では ℓ_1 正則化よりも強く解を疎にするための非凸正則化として対数正則化を用いている. 一方, 本論文では問題の悪条件性を取り除くための凸正則化として対数正則化を用いる. この動機は ℓ_2 正則化を導入する動機と似ているが, 対数正則化は Dirichlet 事前分布や KL 擬距離の罰則化という自然な解釈をもつのに対し, ℓ_2 正則化はそういった解釈を持たない.

さらに本研究では対数正則化を施した問題を解くための近接分離法を 3 つ (加速近接勾配法, 交互方向乗数法, 線形化交互方向乗数法) 提案する. これらのアルゴリズムの各反復で解く子問題はその構造を利用することで効率よく解くことができる.

本論文は次のように構成される: 第 2 節では標準単体上での最小 2 乗問題の応用について述べる. 具体的には混合比率の推定とハイパースペクトル画像の解析の 2 つの応用例を紹介する. 第 3 節では問題 (1) に対する対数正則化を提案し, よく用いられる ℓ_1 正則化や ℓ_2 正則化との比較を行なう. 定理 1 は対数正則化が他の正則化に比べて問題 (1) に適していることを意味している. 第 4 節では対数正則化を施した問題に対する 3 つの近接分離法を提案する. 定理 2–4 は各アルゴリズムの各反復における子問題が効率よく解くことができることを示している.

2 応用例

この節では, われわれの問題, 標準単体上における最小 2 乗問題の 2 つの応用例を示す. 具体的には混合物中における純物質の混合比率の推定とハイパースペクトル画像の解析の 2 つの応用例を紹介する.

2.1 混合比率の推定

問題 (1) は混合物中における純物質の混合比率の推定に用いることができる. 混合物を構成する純物質の物性 (例えば, 質量, 誘電率, 透磁率, 磁化率など) がわかっているものとし, 混合物に関するそれらの物性も測定可能であるものとしよう. ここでは, これらの情報からそれぞれの純物質の混合比率を推定することを考える. 混合物の物性の測定値が各純物質の物性の混合比率による重み付き平均となることを仮定すれば, この混合比の推定問題は問題 (1) として定式化できる. 具体的には, 混合物の物性 i の測定値を b_i, 純物質 j の物性 i の参照値を a_{ij} とし, 残差の 2 乗和を最小化することで純物質 j の混合比率 x_j を推定することができる. 混合比率 $\boldsymbol{x}$ は非負であり, 各比率の総和は 1 となるため, 標準単体制約を課す必要がある. $\boldsymbol{x}$ に対する正則化項 $r(\boldsymbol{x})$ を付け加えると問題 (1) を得る.

Tanaka et al. (2017) は問題 (1) を用いてフォトクロミック分子の混合比率を推定している. フォトクロミック分子とは, ある特定の波長の光を吸収することで化学反応を起こし, それによって変色する分子のことをいう. 彼らは高分子固体中に拡散された反応速度の異なるフォトクロミック分子の混合比を推定するために, その高分子固体に光を照射して透過光の強度の時間変化を測定し, 各時刻における透過光の強度と対応する物性が各分子の反応速度式の解の凸結合になると仮定して, 問題 (1) を用いて混合比率を推定している. 彼らは ℓ_2 正則化を施した問題を商用ソフトウェアを用いて解いている.

2.2 ハイパースペクトル画像の解析

問題 (1) において $r(\boldsymbol{x}) = 0$ としたものはハイパースペクトル画像の解析の文脈でよく研究されている (Bioucas-Dias et al., 2012). ハイパースペクトル画像とは, 狭い波長の幅の電磁波のスペクトルの強さに対応する画像を複数集めた多層の画像のことをいう. ハイパースペクトル画像のそれぞれの画素はその画素に対応する被写体の反射スペクトルに対応する. ハイパースペクトル画像の解析の目的の 1 つは, 端成分と呼ばれる単一の成分のスペクトルとハイパースペクトル画像のある画素のスペクトルからその画素に対応する被写体おける各端成分の存在比率を推定することである. よく用いられる線形モデルでは, 観測されるスペクトルが端成分のスペクトルの凸結合となっていることを仮定する. すなわち, 端成分 j の波長 i の強さ a_{ij} の端成分 j の存在比率 x_j による重み付き平均にノイズ ν_i が加わった $b_i = \sum_{j=1}^{n} a_{ij}x_j + \nu_i$ を観測するものとする. 存在比率は非負かつ総和は 1 であるため, 端成分のスペクトルをあらかじめ知っている場合, 存在比率を推定するためには問題 (1) を解けばよい. 実際, 正則化を施さない問題に対してはいくつかのアルゴリズムが提案されている (Heinz and Chang, 2001; Bioucas-Dias and Figueiredo, 2010; Heylen et al., 2011, 2013; Chouzenoux et al., 2014; Condat, 2017).

ハイパースペクトル画像の解析では, 近い種類の端成分が存在する場合に存在量の推定が不安定になることが知られている (Price, 1994). これは多重共線性の問題に対応する. この問題を回避するために ℓ_2 正則化を導入した研究はいくつか存在する. 黎明期には Settle and Drake (1993) が ℓ_2 正則化を用いた解析を行なっているものの, 制約条件をすべて無視しており, 扱いは十分とはいえない. 近年では Li and Du (2015) が制約条件 $\mathbf{1}^\top \boldsymbol{x} = 1$ を罰則化することで ℓ_2 正則化の扱いをより正確なものにしているが, 非負制約を無視しており, 依然として扱いは十分とはいえない. Chouzenoux et al. (2014, Section 6) は正則化の重要性を認識しており, 正則化を施さない問題に対する彼らのアルゴリズムが正則化を施した問題に容易に拡張できると述べている. しかしながら, 具体的な定式化やアルゴリズムならびに計算機実験の結果は示されていない.

3 対数正則化の性質

この節では対数正則化のいくつかの望ましい性質について述べる. 具体的には, 対数正則化を施した問題は Dirichlet 分布を事前分布とする MAP 推定問題とみなすことができ, それと同時に KL 擬距離を罰則化した問題ともみなすことができることを示す.

3.1 Dirichlet 分布を事前分布とする MAP 推定問題

ここでは, 問題 (1) に対する ℓ_1 正則化, ℓ_2 正則化, 対数正則化を Bayes 的な観点から比較する. 具体的には, 残差 $\nu_i = b_i - \boldsymbol{a}_i^\top \boldsymbol{x}$ がホワイトノイズであるとして問題 (1) と等価な MAP 推定問題を導出し, 各正則化項に対応する事前分布を比較する. 次の定理はそれぞれの正則化項に対応する事前分布を導く.

定理 1. 残差 $\{\nu_i\}_{i=1}^m = \{b_i - \boldsymbol{a}_i^\top \boldsymbol{x}\}_{i=1}^m$ がそれぞれ独立に正規分布 $\mathsf{N}(0, \sigma^2)$ に従うことを仮定し, 問題 (1) における正則化項 r が次のいずれかであるものとする:

1. $r(\boldsymbol{x}) = 0$ (正則化なし),
2. $r(\boldsymbol{x}) = \gamma \|\boldsymbol{x}\|_1$ ($\gamma > 0$, ℓ_1 正則化),
3. $r(\boldsymbol{x}) = (\gamma/2)\|\boldsymbol{x}\|_2^2$ ($\gamma > 0$, ℓ_2 正則化),
4. $r(\boldsymbol{x}) = -\sum_{j=1}^n \gamma_j \log x_j$ ($\boldsymbol{\gamma} > \mathbf{0}$, 対数正則化).

このとき, 問題 (1) はそれぞれ次の事前分布に対応する MAP 推定問題である:

1. $\boldsymbol{x}$ が n 次元標準単体上の一様分布に従う;
2. $\{x_j\}_{j=1}^n$ がそれぞれ独立に Laplace 分布 $\mathsf{Laplace}(0, \sigma^2/\gamma)$ に従う;
3. $\{x_j\}_{j=1}^n$ がそれぞれ独立に正規分布 $\mathsf{N}(0, \sigma^2/\gamma)$ に従う;
4. $\boldsymbol{x}$ が次数 n の Dirichlet 分布 $\mathsf{Dirichlet}(\boldsymbol{\gamma}/\sigma^2 + \mathbf{1})$ に従う.

証明. 問題 (1) は次の最適化問題と等価である:

$$\begin{array}{ll} \text{maximize} & \left(\displaystyle\prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\boldsymbol{a}_i^\top \boldsymbol{x} - b_i)^2}{2\sigma^2}\right)\right) \exp\left(-\frac{r(\boldsymbol{x})}{\sigma^2}\right) \\ \text{subject to} & \mathbf{1}^\top \boldsymbol{x} = 1, \boldsymbol{x} \geq \mathbf{0}. \end{array}$$

目的関数の前半部分 $\prod_{i=1}^m \exp(-(\boldsymbol{a}_i^\top \boldsymbol{x} - b_i)^2/2\sigma^2)/\sqrt{2\pi}\sigma$ は残差 $\{b_i - \boldsymbol{a}_i^\top \boldsymbol{x}\}_{i=1}^m$ の尤度を表す. そのため, 各 r に対して後半部分 $\exp(-r(\boldsymbol{x})/\sigma^2)$ が対応する事前分布の確率密度関数に比例することを示せばよい.

1. $r(\boldsymbol{x}) = 0$ のとき, $\exp(-r(\boldsymbol{x})/\sigma^2) \propto 1/V_n$ が成り立つ. ここで V_n は n 次元標準単体の体積を表す. この式の右辺 $1/V_n$ は n 次元標準単体上の一様分布の確率密度関数に他ならない.
2. $r(\boldsymbol{x}) = \gamma \|\boldsymbol{x}\|_1$ のとき, 次が成り立つ:

$$\exp\left(-\frac{r(\boldsymbol{x})}{\sigma^2}\right) \propto \prod_{j=1}^n \frac{1}{2\sigma^2/\gamma} \exp\left(-\frac{|x_j|}{\sigma^2/\gamma}\right).$$

この右辺は $\{x_j\}_{j=1}^n$ がそれぞれ独立に Laplace 分布 $\mathsf{Laplace}(0, \sigma^2/\gamma)$ に従うときの確率密度関数に一致する.

3. $r(\boldsymbol{x}) = (\gamma/2)\|\boldsymbol{x}\|_2^2$ のとき, 次が成り立つ:

$$\exp\left(-\frac{r(\boldsymbol{x})}{\sigma^2}\right) \propto \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{\gamma}} \exp\left(-\frac{x_j^2}{2\sigma^2/\gamma}\right).$$

この右辺は $\{x_j\}_{j=1}^n$ がそれぞれ独立に正規分布 $\mathsf{N}(0, \sigma^2/\gamma)$ に従うときの確率密度関数に一致する.

4. $r(\boldsymbol{x}) = -\sum_{j=1}^n \gamma_j \log x_j$ のとき, 次が成り立つ:

$$\exp\left(-\frac{r(\boldsymbol{x})}{\sigma^2}\right) \propto \frac{\Gamma(\sum_{j=1}^n \gamma_j/\sigma^2)}{\prod_{j=1}^n \Gamma(\gamma_j/\sigma^2)} \prod_{j=1}^{n} x_j^{\gamma_j/\sigma^2}.$$

ここで Γ はガンマ関数である. この右辺は Dirichlet 分布 $\mathsf{Dirichlet}(\boldsymbol{\gamma}/\sigma^2 + \mathbf{1})$ の確率密度関数に他ならない.

□

この定理は, 正則化を施さない問題と対数正則化を施した問題は自然だが, ℓ_1 正則化や ℓ_2 正則化を施した問題は不自然であることを意味する. 具体的には, 正則化を施さない問題と対数正則化を施した問題に対応する事前分布の台は標準単体またはその内部であって, 問題 (1) の制約条件と一致するのに対し, ℓ_1 正則化や ℓ_2 正則化を施した問題に対応する事前分布の台は $\mathbb{R}^n$ 全体であって, 問題 (1) の制約条件と一致しない. さらに, ℓ_1 正則化や ℓ_2 正則化に対応する事前分布は $\{x_j\}_{j=1}^n$ が独立に同じ分布に従うことや $\mathrm{E}(x_j) = 0$ であることを仮定する. しかしながら, $\boldsymbol{x}$ は $\mathbf{1}^\top \boldsymbol{x} = 1$ と $\boldsymbol{x} \geq \mathbf{0}$ を満たさなければならないため, $\{x_j\}_{j=1}^n$ は独立に同じ分布に従うことはなく, ある j が存在して $\mathrm{E}(x_j) > 0$ である. ℓ_1 正則化や ℓ_2 正則化は事前情報があまりない場合によく用いられる正則化ではあるものの, 今回のように Bayes 的な解釈ができない場合, 対応する統計モデルが意図しないものになっているおそれがある. 一方で, 対数正則化は Bayes 的な解釈をもつため, 対応する統計モデルは明白である. したがって, 十分な事前情報がない下で多重共線性を避けるために正則化を導入する場合, ℓ_2 正則化ではなく対数正則化を導入したほうが安全であると考えられる. もちろん, Dirichlet 事前分布と相反する事前情報をもっている場合は, その事前情報に合った事前分布に対応する正則化を用いたほうがよい.

3.2 KL 擬距離の罰則化

対数正則化は決定変数 $\boldsymbol{x}$ と定数 $\hat{\boldsymbol{x}}$ との間の KL 擬距離の罰則化とみなすこともできる. これはあらかじめ事前情報として $\hat{\boldsymbol{x}}$ があり, それに近い $\boldsymbol{x}$ を求めたい場合に有効である. ℓ_2 損失関数と KL 擬距離 $D(\hat{\boldsymbol{x}}\|\boldsymbol{x}) = \sum_{j=1}^n (\hat{x}_j \log \hat{x}_j - \hat{x}_j \log x_j)$ の和を最小化するためには問題 (1) に $\boldsymbol{\gamma} = \hat{\boldsymbol{x}}$ とした対数正則化を施した問題を解けばよい.

$D(\hat{\boldsymbol{x}}\|\boldsymbol{x})$ の代わりに $\boldsymbol{x}$ と $\hat{\boldsymbol{x}}$ とを入れ替えた $D(\boldsymbol{x}\|\hat{\boldsymbol{x}}) = \sum_{j=1}^n (x_j \log x_j - x_j \log \hat{x}_j)$ を用いることも考えられる. これはエントロピー罰則化 (Koltchinskii, 2009) に対応する. しかしながら, 対数正則化とは違って対応する近接写像の計算を陽に行なうことができないため, この罰則化を導入した問題を解くことは難しい.

KL 擬距離の代わりに ℓ_1 距離 $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_1$ や 2 乗 ℓ_2 距離 $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2$ を用いることもできる. これらは明らかに ℓ_1 正則化や ℓ_2 正則化と関係がある. しかしながら, 例えば 2 乗 ℓ_2 距離と ℓ_2 正則化との間には線形関数の差があるなど, これらは完全に一致するものではない. したがって, ℓ_1 正則化や ℓ_2 正則化は対数正則化のような $\boldsymbol{x}$ と $\hat{\boldsymbol{x}}$ との間の (擬) 距離の罰則化という解釈をもたない.

4 近接分離法

本節では, 問題 (1) に対数正則化 $r(\boldsymbol{x}) = -\sum_{j=1}^{n} \gamma_j \log x_j$ を導入した次の問題を効率よく解く方法を考える:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 - \sum_{j=1}^{n} \gamma_j \log x_j \\ \text{subject to} & \mathbf{1}^\top \boldsymbol{x} = 1, \boldsymbol{x} > \mathbf{0}. \end{array} \tag{2}$$

なお, $x_j \to 0$ のときに目的関数が $+\infty$ に発散することから, ここでは不等式制約 $\boldsymbol{x} \geq \mathbf{0}$ を $\boldsymbol{x} > \mathbf{0}$ に置き換えている. 以下ではこの問題に対する 3 つの近接分離法 (加速近接勾配法, 交互方向乗数法, 線形化交互方向乗数法) を提案する. このうち, 加速近接勾配法には実行可能解の列を生成するという特徴がある. 一方で, 交互方向乗数法と線形化交互方向乗数法には, 主問題と双対問題の実行不能性が容易に計算できるため, 終了条件の設定が容易であるという特徴がある. 加速近接勾配法と交互方向乗数法に関する詳細については, それぞれ Parikh and Boyd (2014); Boyd et al. (2011) やその参考文献を参照されたい.

4.1 加速近接勾配法

加速近接勾配法を適用するための準備として以下の関数を定義する:

$$f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2, \qquad g(\boldsymbol{x}) = -\sum_{j=1}^{n} \gamma_j \log x_j + \iota(\mathbf{1}^\top \boldsymbol{x} = 1) + \iota(\boldsymbol{x} > \mathbf{0}). \tag{3}$$

ここで ι は指示関数である. 関数 f, g は閉真凸関数で, 特に f は微分可能で $\operatorname{dom} f = \mathbb{R}^n$ である. そのため, 問題 (2) と等価な $f(\boldsymbol{x}) + g(\boldsymbol{x})$ の最小化に加速近接勾配法を用いることができる. 加速近接勾配法の擬似コードをアルゴリズム 1 に示す. このアルゴリズムは 2 つのパラメータをもつ. 1 つはステップサイズパラメータ $\lambda^{(k)}$ であり, もう 1 つは加速パラメータ $\omega^{(k)}$ である. ステップサイズパラメータ $\lambda^{(k)}$ は例えば Beck and Teboulle (2009) で用いられているようなバックトラッキングにより定めることができる. また, 加速パラメータ $\omega^{(k)}$ は Nesterov (1983) が提案した $\omega^{(k)} = (\theta^{(k)} - 1)/\theta^{(k+1)}$ を用いることができる. ここで, $\{\theta^{(k)}\}$ は漸化式 $\theta^{(k+1)} = (1/2)(1 + \sqrt{1 + 4(\theta^{(k)})^2})$, $\theta^{(1)} = 1$ で定義される数列である.

アルゴリズム 1 問題 (2) に対する加速近接勾配法

初期点 $\boldsymbol{x}^{(0)}$ とパラメータ $\{\omega^{(k)}\}_{k=1}^{\infty}$ を設定する.
for $k = 0, 1, 2, \ldots,$ (収束するまで):
  $\boldsymbol{y}^{(k+1)} = \boldsymbol{x}^{(k)} + \omega^{(k)}(\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)})$ と更新する.
  パラメータ $\lambda^{(k)}$ を適切に設定する.
  $\boldsymbol{x}^{(k+1)} = \operatorname{prox}_{\lambda^{(k)} g}(\boldsymbol{y}^{(k+1)} - \lambda^{(k)} \nabla f(\boldsymbol{y}^{(k+1)}))$ と更新する.

アルゴリズム 1 における近接写像 $\operatorname{prox}_{\lambda g}(\boldsymbol{v})$ は, 点 $\boldsymbol{v}$ を入力とし, 関数 $g(\boldsymbol{v}) + (1/2\lambda)\|\boldsymbol{x} - \boldsymbol{v}\|_2^2$ を最小にする点を返す写像である. 次の補題と定理はこの最適化問題が半閉形式解をもち, 2 分法を用いてこれを計算することができることを意味する.

補題 1. 正の定数 γ_j, λ と各 $j = 1, \ldots, n$ に対して関数 $x_j(\mu)$ を

$$x_j(\mu) = \frac{1}{2}\left(v_j - \lambda\mu + \sqrt{(v_j - \lambda\mu)^2 + 4\lambda\gamma_j}\right) \tag{4}$$

で定める．このとき，非線形関数 $\sum_{j=1}^n x_j(\mu)$ は連続な単調減少関数である．さらに，非線形方程式 $\sum_{j=1}^n x_j(\mu) = 1$ には唯一の解が存在する．

証明. 連続性については明らかなので，単調性を示す．そのために，各 j について $x_j(\mu)$ が単調であることを示す．いま，

$$\frac{\mathrm{d}x_j}{\mathrm{d}\mu}(\mu) = -\frac{\lambda}{2}\left(1 + \frac{v_j - \lambda\mu}{\sqrt{(v_j - \lambda\mu)^2 + 4\gamma_j\lambda}}\right) = \frac{-2\gamma_j\lambda^2/\sqrt{(v_j - \lambda\mu)^2 + 4\gamma_j\lambda}}{\sqrt{(v_j - \lambda\mu)^2 + 4\gamma_j\lambda} - (v_j - \lambda\mu)}$$

が成り立つ．$\gamma_j > 0$ であるためこの分子は負であり，分子は

$$\sqrt{(v_j - \lambda\mu)^2 + 4\gamma_j\lambda} > \sqrt{(v_j - \lambda\mu)^2} \geq v_j - \lambda\mu$$

であることから正である．以上より，$\mathrm{d}x_j(\mu)/\mathrm{d}\mu < 0$ であるため，$x_j(\mu)$ は単調減少．

連続性と単調性から非線形方程式 $\sum_{j=1}^n x_j^*(\mu) = 1$ の解は存在するとすれば唯一である．ここで，$\mu \to -\infty$ のとき $\sum_{j=1}^n x_j(\mu) \to +\infty$ であることと $\mu \to +\infty$ のとき $\sum_{j=1}^n x_j(\mu) \to 0$ であることを示す．前者は明らかで，後者は次のように示すことができる:

$$\sum_{j=1}^n x_j(\mu) = \sum_{j=1}^n \frac{2\gamma_j\lambda^2}{\sqrt{(v_j - \lambda\mu)^2 + 4\gamma_j\lambda} - (v_j - \lambda\mu)} \to 0.$$

したがって，中間値の定理から非線形方程式 $\sum_{j=1}^n x_j^*(\mu) = 1$ の解が存在することがわかる． □

定理 2. 関数 g と x_j をそれぞれ式 (3) と式 (4) で定義されるものとする．このとき，近接写像 $\mathrm{prox}_{\lambda g}(\boldsymbol{v})$ は半閉形式解 $(\mathrm{prox}_{\lambda g}(\boldsymbol{v}))_j = x_j(\mu^*)$ $(j = 1, \ldots, n)$ をもつ．ここで μ^* は非線形方程式 $\sum_{j=1}^n x_j(\mu) = 1$ の唯一の解である．

証明. 点 $\boldsymbol{v}$ における近接写像の値 $\mathrm{prox}_{\lambda g}(\boldsymbol{v})$ は次の最適化問題の解である:

$$\begin{aligned} &\text{minimize} && -\sum_{j=1}^n \gamma_j \log x_j + \frac{1}{2\lambda}\|\boldsymbol{x} - \boldsymbol{v}\|_2^2 \\ &\text{subject to} && \mathbf{1}^\top \boldsymbol{x} = 1, \boldsymbol{x} > \mathbf{0}. \end{aligned} \tag{5}$$

この最適化問題において不等式制約を無視し，等式制約に対する Lagrange 乗数を μ したときの最適性条件は，$\sum_{j=1}^n x_j = 1$ と同時に各 $j = 1, \ldots, n$ に対して

$$x_j^2 - (v_j - \lambda\mu)x_j - \lambda\gamma_j = 0 \tag{6}$$

を満たすことである．$\lambda\gamma_j > 0$ であるため，式 (6) は常に唯一の正の根を持つ．各 j に対してその正の根をとると式 (4) を得る．また，μ^* は補題 1 からその存在が保証され，$(x_1(\mu^*), \ldots, x_n(\mu^*))$ が問題 (5) に対する最適性条件を満たすことがわかる． □

4.2 交互方向乗数法とその変種

問題 (2) を加速近接勾配法のときとは異なる 2 つの関数に分離させることにより, 以下で述べるように子問題がより簡単に解けるようになり, 2 分法を行なう必要がなくなる. しかしながら, 以下で述べる分離の方法では定義域全体での微分可能性がなくなるため, 加速近接勾配法ではなく交互方向乗数法やその亜種を用いる必要がある.

4.2.1 交互方向乗数法

以下では問題 (2) を次の 2 つの関数の和の最小化に分離する:

$$h(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \iota(\mathbf{1}^\top \boldsymbol{x} = 1), \qquad l(\boldsymbol{x}) = -\sum_{j=1}^{n} \gamma_j \log x_j + \iota(\boldsymbol{x} > \mathbf{0}). \tag{7}$$

関数 h, l は閉真凸関数なので, 問題 (2) と等価な $h(\boldsymbol{x}) + l(\boldsymbol{x})$ の最小化に交互方向乗数法を適用することができる. 交互方向乗数法の擬似コードをアルゴリズム 2 に示す. このアルゴリズムはステップサイズパラメータ $\lambda^{(k)}$ をもつ. この値は He et al. (2000); Wang and Liao (2001) が提案した方法を用いて適応的に決めることができる. また, $\|\boldsymbol{x}^{(k)} - \boldsymbol{z}^{(k)}\|_2$ は主問題の実行不能性, $\|\boldsymbol{z}^{(k)} - \boldsymbol{z}^{(k-1)}\|_2$ は双対問題の実行不能性に対応するため, これらがともに十分小さくなったときにアルゴリズムを終了すれば, そのときの解は最適解に十分近いといえる.

アルゴリズム 2 問題 (2) に対する交互方向乗数法

初期点 $\boldsymbol{z}^{(0)}$ と初期パラメータ $\lambda^{(0)}$ を適当に決める.
for $k = 0, 1, 2, \ldots,$ (収束するまで):
  $\boldsymbol{x}^{(k+1)} = \mathrm{prox}_{\lambda^{(k)} h}(\boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)})$ と更新する.
  $\boldsymbol{z}^{(k+1)} = \mathrm{prox}_{\lambda^{(k)} l}(\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)})$ と更新する.
  $\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k+1)}$ と更新する.
  必要があればパラメータ $\lambda^{(k+1)}$ を更新し, $\boldsymbol{u}^{(k+1)}$, を適当にスケーリングする.

アルゴリズム 2 に現れる近接写像 $\mathrm{prox}_{\lambda h}(\boldsymbol{v}), \mathrm{prox}_{\lambda l}(\boldsymbol{v})$ は陽に計算することができる. 実際, l の近接写像は $(\mathrm{prox}_{\lambda l}(\boldsymbol{v}))_j = (1/2)(v_j + \sqrt{v_j^2 + 4\lambda\gamma_j})$ $(j = 1, \ldots, n)$ であることがよく知られている. 次の定理は $\mathrm{prox}_{\lambda h}(\boldsymbol{v})$ の計算式を陽に与えるものである.

定理 3. 関数 h を式 (7) で定められるものとする. このとき, 近接写像 $\mathrm{prox}_{\lambda h}(\boldsymbol{v})$ は次のようにかける:

$$\mathrm{prox}_{\lambda h}(\boldsymbol{v}) = \left(\boldsymbol{A}^\top \boldsymbol{A} + \frac{1}{\lambda}\boldsymbol{I}\right)^{-1} \left(\boldsymbol{A}^\top \boldsymbol{b} + \frac{1}{\lambda}\boldsymbol{v} - \mu \mathbf{1}\right). \tag{8}$$

ここで,

$$\mu = \frac{\mathbf{1}^\top (\boldsymbol{A}^\top \boldsymbol{A} + (1/\lambda)\boldsymbol{I})^{-1} (\boldsymbol{A}^\top \boldsymbol{b} + (1/\lambda)\boldsymbol{v}) - 1}{\mathbf{1}^\top (\boldsymbol{A}^\top \boldsymbol{A} + (1/\lambda)\boldsymbol{I})^{-1} \mathbf{1}} \tag{9}$$

である.

証明. 近接写像 $\mathrm{prox}_{\lambda h}(\boldsymbol{v})$ は次の最適化問題の最適解である:

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+\frac{1}{2\lambda}\|\boldsymbol{x}-\boldsymbol{v}\|_2^2\\ \text{subject to} & \boldsymbol{1}^\top\boldsymbol{x}=1.\end{array}$$

この問題に対する最適性条件は μ を Lagrange 乗数として

$$\left(\boldsymbol{A}^\top\boldsymbol{A}+\frac{1}{\lambda}\boldsymbol{I}\right)\boldsymbol{x}+\mu\boldsymbol{1}=\boldsymbol{A}^\top\boldsymbol{b}+\frac{1}{\lambda}\boldsymbol{v},$$
$$\boldsymbol{1}^\top\boldsymbol{x}=1$$

である. 上の方程式を $\boldsymbol{x}$ について解くと次を得る:

$$\boldsymbol{x}=\left(\boldsymbol{A}^\top\boldsymbol{A}+\frac{1}{\lambda}\boldsymbol{I}\right)^{-1}\left(\boldsymbol{A}^\top\boldsymbol{b}+\frac{1}{\lambda}\boldsymbol{v}-\mu\boldsymbol{1}\right).$$

この右辺は式 (8) の右辺に他ならない. さらに, これを下の方程式に代入すると次を得る:

$$\boldsymbol{1}^\top\left(\boldsymbol{A}^\top\boldsymbol{A}+\frac{1}{\lambda}\boldsymbol{I}\right)^{-1}\left(\boldsymbol{A}^\top\boldsymbol{b}+\frac{1}{\lambda}\boldsymbol{v}\right)-\mu\boldsymbol{1}^\top\left(\boldsymbol{A}^\top\boldsymbol{A}+\frac{1}{\lambda}\boldsymbol{I}\right)^{-1}\boldsymbol{1}=1.$$

この方程式を μ について解くと, 式 (9) を得る. □

注意 1. 式 (8)–(9) における $(\boldsymbol{A}^\top\boldsymbol{A}+(1/\lambda)\boldsymbol{I})^{-1}$ の計算は $\boldsymbol{A}$ の特異値分解 $\boldsymbol{U}\,\mathrm{Diag}(\sigma_1,\dots,\sigma_p)\boldsymbol{V}^\top$ を用いれば効率よく行なうことができる. ここで $p=\min\{m,n\}$ である. 実際, $m<n$ について $\sigma_{m+1}=\cdots=\sigma_n=0$ と定めると

$$\left(\boldsymbol{A}^\top\boldsymbol{A}+\frac{1}{\lambda}\boldsymbol{I}\right)^{-1}=\boldsymbol{V}\,\mathrm{Diag}\left(\frac{\lambda}{\lambda\sigma_1^2+1},\dots,\frac{\lambda}{\lambda\sigma_n^2+1}\right)\boldsymbol{V}^\top$$

である. したがって, あらかじめ $\boldsymbol{A}$ の特異値分解を計算し, $\sigma_1,\dots,\sigma_p$ と $\boldsymbol{V}$ を記憶しておけば, 任意の λ に対して $(\boldsymbol{A}^\top\boldsymbol{A}+(1/\lambda)\boldsymbol{I})^{-1}$ の計算を効率よく行なうことができる.

4.2.2 線形化交互方向乗数法

前述の通り, 交互方向乗数法では $\boldsymbol{A}$ の特異値分解をあらかじめ計算しておけば $(\boldsymbol{A}^\top\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}$ の計算を高速に行なうことができる. しかしながら, 特異値分解の計算量は $O(\min\{m^2n,mn^2\})$ とみなすことができ, 大規模な問題例に対しては非常に大きくなる. そこで, $(\boldsymbol{A}^\top\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}$ の計算を避けるために, 不正確な Uzawa 法 (Zhang et al., 2010a,b) としても知られている線形化交互方向乗数法を用いることを考える. このアルゴリズムの各反復では次のように $\boldsymbol{x}$ を更新する:

$$\boldsymbol{x}^{(k+1)}=\operatorname*{argmin}_{\boldsymbol{x}}\left\{h(\boldsymbol{x})+\frac{1}{2\lambda^{(k)}}\|\boldsymbol{x}-\boldsymbol{z}^{(k)}+\boldsymbol{u}^{(k)}\|_2^2+\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}^{(k)}\|_{\boldsymbol{G}}^2:\boldsymbol{1}^\top\boldsymbol{x}=1\right\}. \tag{10}$$

ここで $\|\boldsymbol{x}\|_{\boldsymbol{G}}=\sqrt{\boldsymbol{x}^\top\boldsymbol{G}\boldsymbol{x}}$ は半正定値対称行列 $\boldsymbol{G}$ によって定まるノルムである. 次の定理は $\boldsymbol{G}$ をうまく取れば子問題 (10) を非常に簡単に解くことができることを示している.

定理 4. 実数 α を $\alpha\geq\sigma_{\max}(\boldsymbol{A})^2$ 満たすようにとり, $\boldsymbol{G}=\alpha\boldsymbol{I}-\boldsymbol{A}^\top\boldsymbol{A}$ とする. このとき, $\boldsymbol{G}$ は半正定値である. さらに, 子問題 (10) は次の閉形式の解をもつ:

$$\boldsymbol{x}^{(k+1)}=\frac{\lambda^{(k)}}{\alpha\lambda^{(k)}+1}(\boldsymbol{r}-\mu\boldsymbol{1}). \tag{11}$$

ここで $\boldsymbol{r} = \boldsymbol{A}^\top \boldsymbol{b} + (1/\lambda^{(k)})(\boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)}) + \alpha \boldsymbol{x}^{(k)} - \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}^{(k)}$ であり,

$$\mu = \frac{1}{n}\left(\mathbf{1}^\top \boldsymbol{r} - \frac{1}{\lambda^{(k)}} - \alpha\right) \tag{12}$$

である.

証明. $\|\boldsymbol{v}\|_2 = 1$ であるような任意の $\boldsymbol{v}$ に対し, 次が成り立つ:

$$\boldsymbol{v}^\top \boldsymbol{G}\boldsymbol{v} = \alpha - \boldsymbol{v}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{v} = \alpha - \|\boldsymbol{A}\boldsymbol{v}\|_2^2 \geq \alpha - \sigma_{\max}(\boldsymbol{A})^2 \geq 0.$$

したがって, $\boldsymbol{G}$ は半正定値. さらに, 子問題 (10) は次の問題と等価である:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\left(\frac{1}{\lambda^{(k)}} + \alpha\right)\boldsymbol{x}^\top \boldsymbol{x} - \boldsymbol{r}^\top \boldsymbol{x} \\ \text{subject to} & \mathbf{1}^\top \boldsymbol{x} = 1. \end{array}$$

この問題に対する最適性条件は式 (11)–(12) に他ならない. □

4.3 理論的な比較

加速近接勾配法の最良の収束率が $O(1/k^2)$ であることは Beck and Teboulle (2009, Theorem 4.4) によって示されている. また, f が強凸のときに線形収束することも Schmidt et al. (2011, Proposition 4) によって示されている. 一方で, 交互方向乗数法と線形化交互方向乗数法の現時点での最良の収束率は $O(1/k)$ である (He and Yuan, 2012, Theorem 4.1). これらのアルゴリズムの実際の収束率が $O(1/k)$ であるとすると, 加速近接勾配法は他の 2 つのアルゴリズムと比較して少ない反復回数で収束することが期待できる.

一方で, 交互方向乗数法と線形化交互方向乗数法では加速近接勾配法における直線探索や 2 分法を行なう必要がない. そのため, 各反復での計算は交互方向乗数法と線形化交互方向乗数法の方が高速に行なうことができる. さらに, 線形化交互方向乗数法では $\boldsymbol{A}$ の特異値分解さえする必要がない. したがって, 線形化交互方向乗数法は交互方向乗数法に比べて大規模な問題に対しては有効であることが期待される.

参考文献

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problem. *SIAM Journal on Imaging Sciences*, 2:183–202.

Bioucas-Dias, J. M. and Figueiredo, M. A. T. (2010). Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In *The Proceedings of The 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4.

Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5:354–379.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122.

Candés, E. J., Wakin, M. B., and Boyd, S. P. (2008). Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905.

Chouzenoux, E., Legendre, M., Moussaoui, S., and Idier, J. (2014). Fast constrained least squares spectral unmixing using primal-dual interior point optimization. *IEEE Transactions on Geoscience and Remote Sensing*, 7:59–69.

Condat, L. (2017). Least-squares on the simplex for multispectral unmixing. Technical Report, GIPSA-Lab, Grenoble, France.

Dai, Y.-H. and Fletcher, R. (2006). New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Mathematical Programming*, 106:403–421.

Han, C., Li, M., Zhao, T., and Guo, T. (2013). An accelerated proximal gradient algorithm for singly linearly constrained quadratic programs with box constraints. *The Scientific World Journal*, 2013(246596).

He, B. and Yuan, X. (2012). On the $o(1/n)$ convergence rate of the douglas–rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50:700–709.

He, B.-S., Yang, H., and Wang, S. L. (2000). Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106:337–356.

Heinz, D. C. and Chang, C.-I. (2001). Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39:529–545.

Heylen, R., Akhter, M. A., and Scheunders, P. (2013). Solving the hyperspectral unmixing problem with projection onto convex sets. In *The Proceedings of The 21st European Signal Processing Conference*, pages 1–5.

Heylen, R., Burazerovic, D., and Scheunders, P. (2011). Fully constrained least-squares spectral unmixing by simplex projection. *IEEE Transactions on Geoscience and Remote Sensing*, 49:4112–4122.

Koltchinskii, V. (2009). Sparse recovery in convex hulls via entropy penalization. *The Annals of Statistics*, 37:1332–1359.

Larsson, M. O. and Ugander, J. (2011). A concave regularization technique for sparse mixture models. In *Advances in Neural Information Processing Systems 24*, pages 1890–1898.

Li, W. and Du, Q. (2015). Collaborative representation for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 53:1463–1474.

Mazumder, R., Friedman, J. H., and Hastie, T. (2011). SparseNet: coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, 106:1125–1138.

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Soviet Mathematics Doklady*, 27:372–376.

Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in Optimization*, 1:123–231.

Price, J. C. (1994). How unique are spectral signatures? *Remote Sensing of Environment*, 49:181–186.

Schmidt, M., Roux, N. L., and Bach, F. R. (2011). Convergence rates of inexact proximal-gradient methods for convex optimization. In *Advances in Neural Information Processing Systems 24*, pages

1458–1466.

Settle, J. J. and Drake, N. A. (1993). Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, 14:1159–1177.

Tanaka, M., Yamashita, T., Sano, N., Ishigaki, A., and Suzuki, T. (2017). Mathematical optimization approach for estimating the quantum yield distribution of a photochromic reaction in a polymer. *AIP Advances*, 7(015041).

Tibshirani, R. (1994). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58:267–288.

Wang, S. L. and Liao, L.-Z. (2001). Decomposition method with a variable parameter for a class of monotone variational inequality problems. *Journal of Optimization Theory and Applications*, 109:415–429.

Weston, J., Elisseeff, A., Schölkopf, B., and Tipping, M. (2003). Use of the zero-norm with linear models and kernel methods. *Journal of Machine Learning Research*, 3:1439–1461.

Zhang, X., Burger, M., Bresson, X., and Osher, S. (2010a). Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3:253–276.

Zhang, X., Burger, M., and Osher, S. (2010b). A unified primal-dual algorithm framework based on bregman iteration. *SIAM Journal on Scientific Computing*, 46:20–46.